

Завдання для виконання лабораторних робіт.

Практичне завдання.

1. Створити аккаунт на Github.
2. Створити новий репозиторій під назвою DataScienceCourseKNU
3. Розмістити в репозиторії 2 файли: readme.md – з заголовком та коротким описом репозиторію та HelloWorld.md

Інструкції для виконання лабораторних робіт.

Для виконання лабораторних робіт необхідно встановити R та R Studio.

Посилання для скачування програм:

1. R: «<https://cran.r-project.org/>»
2. R Studio: «<https://www.rstudio.com/products/rstudio/download/>»

Результати лабораторних робіт необхідно розмістити в своєму репозиторії на Github у файлі «LabX.md», де X – номер лабораторної у вигляді: питання, код, результат (якщо треба). Наприклад:

1. Які назви строк фрейму даних?

```
rownames(dataset)
```

```
## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11"
## [12] "12" "13" "14" "15" "16" "17" "18" "19" "20" "21" "22"
## [23] "23" "24" "25" "26" "27" "28" "29" "30" "31" "32" "33"
## [34] "34" "35" "36" "37" "38" "39" "40" "41" "42" "43" "44"
## [45] "45" "46" "47" "48" "49" "50" "51" "52" "53" "54" "55"
## [56] "56" "57" "58" "59" "60" "61" "62" "63" "64" "65" "66"
## [67] "67" "68" "69" "70" "71" "72" "73" "74" "75" "76" "77"
## [78] "78" "79" "80" "81" "82" "83" "84" "85" "86" "87" "88"
## [89] "89" "90" "91" "92" "93" "94" "95" "96" "97" "98" "99"
## [100] "100" "101" "102" "103" "104" "105" "106" "107" "108" "109" "110"
## [111] "111" "112" "113" "114" "115" "116" "117" "118" "119" "120" "121"
## [122] "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143"
## [144] "144" "145" "146" "147" "148" "149" "150" "151" "152" "153"
```

Лабораторна робота № 1

В лабораторній роботі необхідно виконати наступні дії:

1. Створити змінні базових (atomic) типів. Базові типи: character, numeric, integer, complex, logical.
2. Створити вектори, які: містить послідовність з 5 до 75; містить числа 3.14, 2.71, 0, 13; 100 значень TRUE.
3. Створити наступну матрицю за допомогою **matrix**, та за допомогою **cbind**

або `rbind`

0.5	1.3	3.5
3.9	131	2.8
0	2.2	4.6
2	7	5.1

4. Створити довільний список (list), в який включити всі базові типи.
5. Створити фактор з трьома рівнями «baby», «child», «adult».
6. Знайти індекс першого значення NA в векторі 1, 2, 3, 4, NA, 6, 7, NA, 9, NA, 11. Знайти кількість значень NA.
7. Створити довільний data frame та вивести в консоль.
8. Змінити імена стовпців цього data frame.

Лабораторна робота № 2

В лабораторній роботі необхідно виконати наступні дії:

1. Створити вектор `v` із 100 елементів командою `v <- rnorm(100)`. Для цього вектору виведіть: 10-й елемент; елементи з 10-го по 20-й включно; 10 елементів починаючи з 20-го; елементи більше 0.
2. Створити фрейм (data frame) `y` командою `y <- data.frame(a = rnorm(100), b = 1:100, ss = sample(letters, 100, replace = TRUE))`. Для цього data frame виведіть: останні 10 строк; строки з 10 по 20 включно; 10-й елемент стовпця `b`; повністю стовпець `ss`, при цьому використовуйте ім'я стовпця.
3. Створити вектор `z` з елементами 1, 2, 3, NA, 4, NA, 5, NA. Для цього вектору: виведіть всі елементи, які не NA; підрахуйте середнє значення всіх елементів цього вектору без NA значень та з NA значеннями.

Лабораторна робота № 3

В лабораторній роботі необхідно написати наступні функції на мові R та вивести результат роботи цих функцій на довільних даних:

1. Функція `add2(x, y)`, яка повертає суму двох чисел.
2. Функція `above(x, n)`, яка приймає вектор та число `n`, та повертає всі елементи вектору, які більше `n`. По замовчуванню `n = 10`.
3. Функція `my_ifelse(x, expr, n)`, яка приймає вектор `x`, порівнює всі його елементи за допомогою `expr` з `n`, та повертає елементи вектору, які відповідають умові `expression`. Наприклад, `my_ifelse(x, ">", 0)` повертає всі елементи `x`, які більші 0. `Expr` може дорівнювати "`<`", "`>`", "`<=`", "`>=`", "`==`". Якщо `expr` не співпадає ні з одним з цих виразів, повертається вектор `x`.
4. Функція `columnmean(x, removeNA)`, яка розраховує середнє значення (mean) по кожному стовпцю матриці, або data frame. Логічний параметр `removeNA` вказує, чи видаляти NA значення. По замовчуванню він дорівнює TRUE.

Лабораторна робота № 4

Для лабораторної роботи необхідно завантажити файл з даними за посиланням: «https://www.dropbox.com/s/8k1gigk8cflmpb6/hw1_data.csv?dl=0». В RStudio файл необхідно прочитати за допомогою команди `read.csv`.

В лабораторній необхідно відповісти на запитання:

1. Які назви стовпців файлу даних?
2. Виведіть перші 6 строк фрейму даних.
3. Скільки спостережень (строк) в дата фреймі?
4. Виведіть останні 10 строк дата фрейму.
5. Як багато значень «NA» в стовпці «Ozone»?
6. Яке середнє (mean) стовпця «Ozone». Виключити «NA» значення.
7. Виведіть частину набору даних (subset) зі значенням «Ozone» > 31 та «Temp» > 90. Яке середнє (mean) значень «Solar.R» в цьому наборі даних (subset)?
8. Яке середнє значення (mean) для «Temp» для червня («Month» дорівнює 6)?
9. Яке максимальне значення «Ozone» для травня («Month» дорівнює 5)?

Лабораторна робота № 5

Для лабораторної роботи необхідно завантажити zip файл з даними за посиланням:

«https://www.dropbox.com/s/i9wi47oyhfb7qlh/rprog_data_specdata.zip?dl=0».

Цей файл містить 332 csv файлів, що містять у собі результати спостережень за забрудненням повітря дрібнодисперсними частками (fine particulate matter air pollution) у 332 локаціях у США. Кожен файл містить дані з одного монітору. ID номер кожного монітору міститься у назві файлу. Наприклад, дані з монітору під номером 200 містяться у файлі «200.csv». Кожен файл містить три змінні:

Data: дата спостереження в форматі (рік-місяць-день), **sulfate:** рівень сульфатних часток в повітрі на дату (мікрограми на кубічний метр) та **nitrate:** рівень нітратних часток в повітрі на дату (мікрограми на кубічний метр). Для цієї роботи необхідно додати на Github файл `rmean.R`, який містить усі функції. В файлі `md` необхідно указати виклик функції з аргументами та вивід у консоль результатів роботи функцій.

1. Написати функцію `rmean`, яка обчислює середнє значення (mean) забруднення сульфатами або нітратами серед заданого переліка моніторів. Ця функція приймає три аргументи: «directory», «pollutant», «id». Directory – папка, в якій розміщені дані, pollutant – вид забруднення, id – перелік моніторів. Аргумент id має значення за замовчуванням 1:332. Функція повинна ігнорувати NA значення. Приклад роботи функції:

```
rmean("specdata", "sulfate", 1:10)
## [1] 4.064128
rmean("specdata", "sulfate", 55)
```

```
## [1] 3.587319
pmean("specdata", "nitrate")
## [1] 1.702932
```

2. Написати функцію `complete`, яка виводить кількість повних спостережень (the number of completely observed cases) для кожного файлу. Функція приймає два аргументи: «Directory» та «id» та повертає data frame, в якому перший стовпчик – ім'я файлу, а другий – кількість повних спостережень. Приклад роботи функції:

```
complete("specdata", 1)
## id nobs
## 1 1 117

complete("specdata", c(2, 4, 8, 10, 12))
## id nobs
## 1 2 1041
## 2 4 474
## 3 8 192
## 4 10 148
## 5 12 96
```

```
complete("specdata", 50:60)
## id nobs
## 1 50 459
## 2 51 193
## 3 52 812
## 4 53 342
## 5 54 219
## 6 55 372
## 7 56 642
## 8 57 452
## 9 58 391
## 10 59 445
## 11 60 448
```

3. Написати функцію `corr`, яка приймає два аргументи: `directory` (папка, де знаходяться файли спостережень) та `threshold` (порогове значення, за замовчуванням дорівнює 0) та обчислює кореляцію між сульфатами та нітратами для моніторів, кількість повних спостережень для яких більше порогового значення. Функція повинна повернути вектор значень кореляцій. Якщо ні один монітор не перевищує порогового значення, функція повинна повернути numeric вектор довжиною 0. Для обчислення кореляції між сульфатами та нітратами використовуйте вбудовану функцію

«corr» з параметрами за замовчуванням.

Приклад роботи функції:

```
cr <- corr("specdata", 150)
```

```
head(cr); summary(cr)
```

```
## [1] -0.01895754 -0.14051254 -0.04389737 -0.06815956 -0.12350667 -  
0.07588814
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
```

```
## -0.21060 -0.04999 0.09463 0.12530 0.26840 0.76310
```

```
cr <- corr("specdata", 400)
```

```
head(cr); summary(cr)
```

```
## [1] -0.01895754 -0.04389737 -0.06815956 -0.07588814 0.76312884 -  
0.15782860
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
```

```
## -0.17620 -0.03109 0.10020 0.13970 0.26850 0.76310
```

```
cr <- corr("specdata", 5000)
```

```
head(cr); summary(cr) ; length(cr)
```

```
## NULL
```

```
## Length Class  Mode
```

```
##    0  NULL  NULL
```

```
## [1] 0
```